

The value of safe settings in evidence based policy making

Sarah Lowe

Lead, Programme to Maximise the Use of Existing Data, Welsh Government sarah.lowe@gov.wales

Abstract

This paper will look at the value that a safe setting infrastructure can offer to the UK government and public. It will use examples from the Welsh Government's Programme to Maximise the Use of Existing Data and international organisations to demonstrate the value that can be added by enabling secure linked data to be processed and accessed via infrastructures such as the ADRN. The paper will focus particularly on evidence based policy making and the reassurance that safe settings can offer the British public drawing on research by IPSOS-Mori, the Royal Statistical Society and the Law Commission

Keywords – Higher Education; Homelessness Prevention; Suicide; Fuel Poverty; Children in Care

1 Introduction

Central and Local Government Departments and Agencies collect a significant amount of routine information about individual citizens as part of their everyday activities. Significant investment is also made by the Public Sector in data collection for social survey and other primary research and evaluation projects.

Government agencies both across the UK and Internationally are becoming increasingly interested in data sharing to make better use of existing information resources for evidence-informed policy making and service delivery. In hard economic times, the improved use of existing data can be an important way of adding value, advancing the *Collect once, Use many times* paradigm. As the other papersⁱ given for this session demonstrate, the use of safe settings can provide a secure infrastructure, giving government the confidence to share the kinds of individual level data necessary for linking so that we can make the best use of that data.

The volume and richness of 'administrative data' can vastly exceed what it's possible to collect through other routes such as surveys. However, sheer volume can be a double-edged sword since, where the data are not designed for analytical purposes, researchers often encounter uneven data quality and completeness, challenges for the

uninitiated in navigating complexity and understanding content, and significant challenges in combining data to define conditions e.g. diabetes, or concepts e.g. socio-economic status, using administrative data.

As a government researcher of more than 20 years, I'm interested in data linking because, among other benefits, it can help me solve two problems I've been wrestling with throughout my career. The first of these problems is that policymakers often don't include analysts at an early enough stage in their plans so we can be left without a baseline to evaluate against – due to its longitudinal nature, linked data can offer the possibility of constructing a baseline retrospectively. The second problem is the disconnect between policy and research timescales. Understandably, policymakers and ministers need evidence as quickly as possible to inform their decision making, whilst analysts are often left having to explain that a good quality survey with a suitable sample size or a robust evaluation study will cost a lot of money and take several years to report. If we get both our secure infrastructure and our routine flows of administrative data right, in the long run, I believe we can begin to narrow this gap.

I'm not saying it's going to happen tomorrow. I began working in the field of data linking in 2009, at which point I would have confidently predicted I could have achieved all the ambitious visions I had for improving the evidence base in three to five years. Eight years later, and with a much better understanding of the challenges of working with administrative data, I'd say that if we in Government invest both the time and the money now, in perhaps another ten years we could have an analytical infrastructure that would be the envy of the world. And I don't just mean a range of secure settings through which datasets can be securely and anonymously linked for research purposes. I mean a range of administrative datasets flowing routinely from data management systems into secure settings and out again, as aggregates, into dashboards, desktop data portals, publically accessible interactive data websites like [StatsWales](#)ⁱⁱ and [Lle](#)ⁱⁱⁱ and a variety of semi-automated statistical bulletins and research, monitoring and evaluation reports, freeing up analytical time for more detailed interpretation and to draw out the implications of research findings for our policy colleagues.

2. But we need to get it right!

Don't misunderstand me, there are certainly huge challenges to be overcome along the way. As noted above, there are significant challenges with the use of administrative data springing from the fact it was not collected for research purposes. These problems range from identifiable and sensitive information being found in unexpected places (and having to be stripped out using Natural Language Processing software tools, the quality of which is improving all the time) to key information being missing or recorded inconsistently by service delivery professionals or provider organisations.

Feeding back information about data quality to allow the improvement of upstream data collection needs to be a significant part of many data linking projects as they begin working with providers of administrative data. Where data providers are almost always trying to find the time to assist us on top of their day job, careful consideration needs to be given to both 'carrots' and 'sticks' when trying to acquire data, ensure the associated metadata is comprehensive and improve data collection. For example, Welsh Government is starting to include the requirement to provide individual-level data for anonymisation into the SAIL^{iv} secure setting as part of the grant Terms and Conditions for major social policy programmes; however, we are also part-funding jointly with the ESRC a pilot project to install National Research Data Appliances in five local authorities in Wales – among other benefits, these 'linking engines' will allow local authorities to develop regular routine flows of anonymised administrative data from their data management systems to SAIL with, after the initial set-up time, the minimum ongoing resource requirements.

We also need to make sure to take both the various publics and the human rights and privacy organisations with us. Research by organisations such as IPSOS Mori on behalf of the ONS and ESRC^v, The Wellcome Trust^{vi}, Royal Statistical Society^{vii}, and the Law Commission all shows that the better the public understand how we securely process and store their data and the uses to which it's put, the more comfortable they are with anonymisation and linking. In particular, the public are reassured by the kinds of safeguards safe settings can provide: secure data storage, strict access controls, data destruction protocols, and transparency around outputs and publications. As Allison Park, the Director of the CLOSER^{viii} programme points out, attention must *"rightly be paid to the ethical considerations that govern who has access to this kind of personal information and under what conditions. It is vital that the research community engages with these debates and provides compelling examples of how and why we use admin data, and the public benefit this generates"*.^{ix}

When scandals like Care.data^x can easily shatter public confidence, we need to be both measured and inclusive as we develop and implement strategies for sharing the personal, potentially sensitive routine information we, as

government, hold about individual citizens. As Professor Fiona Stanley of the University of Western Australia has argued, we need to get the message out that *'we're not interested in individual people; we're interested in large numbers and [in] link[ing] that to outcomes on health, on education, on child protection, on public housing'*^{xi}.

3. The bits only Geeks really care about...

There are numerous advantages to the use of linked data that may not directly assist you as policymakers but which your analytical colleagues will certainly thank you for. However, in the long run, these more technical developments can have an impact on analytical resources, budgets and delivery timescales.

For example, if survey respondents give their consent, significant added value can be gained from linking the rich, self-reported information collected in social surveys with the detail that can be gleaned from the administrative data held about the same people. Due to plummeting response rates and survey budgets, we're rarely able to collect the amount of detail we'd ideally like to collect using surveys but by linking to administrative data we can provide supplementary information e.g. on health conditions or receipt of benefits, shorten interview length or allow factual questions to be replaced with behavioural or attitudinal questions, potentially saving money and/or increasing response rates.

For some topics, it may be possible to provide indicative estimates using existing administrative data brought together through linking, reducing the amount of expensive new data collection required or targeting it more effectively.

From a methodological point of view, linking survey data to administrative sources can allow comparisons to be made between self-reported information and information available from public services. This can be used to triangulate between sources, investigate mode effects and enhance our understanding of both kinds of data. Linking to administrative data can also help us to understand the nature of non-response and to construct more accurate weights to control for it. So, whilst each type of data has its strengths and weaknesses, the evidence base can be significantly improved by bringing the two together.

4. The Opportunities for Policy

In addition to the methodological advantages discussed above, there are a range of benefits that have a more direct impact on the evidence base.

As noted above, it is possible to construct baselines retrospectively e.g. examining vaccination rates before the Welsh Government's *Flying Start* programme started providing enhanced health visiting to the parents of deprived children aged under 4 years in Wales, but we can

also identify suitable ‘control’ cases using the more detailed kinds of information we can link to about the characteristics of the rest of the population. Once flagged, and without the need for the kinds of expensive ‘keeping in touch exercises’ that would be required for a survey, we can follow our anonymised ‘intervention’ and ‘control’ cohorts over time to observe their long-term outcomes e.g. to unpick whether our homelessness prevention programme, *Supporting People* gives recipients long-term improvements in health, housing, labour market and child educational attainment outcomes compared with controls. We are currently evaluating the feasibility of building a ‘cost-offsetting’ model within the SAIL safe setting, fed by *Supporting People* administrative data and calculating ‘adverse events’ for recipients and controls using e.g. health events from routine health records, allowing a more accurate estimate to be made of any savings to the NHS in Wales.

In fact, using the richer, more crosscutting datasets that can be linked in a secure setting, the accuracy of a range of statistical models used by government can potentially be improved. For example, whilst Welsh Government is currently undertaking our most recent survey of house conditions, in parallel we are developing a *Housing Stock Analytical Resource (HSAR Wales/Cymru)* by anonymously and securely linking information from all available sources relating to domestic dwellings in Wales. Not only will *HSAR* provide an ongoing resource allowing us to answer a range of research and statistical questions about house conditions and fuel poverty in Wales but where small area estimates are constructed based on modelled house condition survey data, we expect the robustness of the associated estimates to be significantly improved due to being able to take into account more detailed information about our wider housing stock.

Moreover, because administrative data is by its very nature longitudinal, it can allow cause and effect to be disentangled, for example, as in a soon-to-be-completed Welsh Government-ESRC-funded PhD^{xii}, the extent to which disability may drive issues around labour market participation. Longitudinal data allows us to examine the dynamics of issues such as poverty and social mobility. By securely linking anonymised data, we can, for example, begin to answer questions such as whether the Welsh Government’s *Warm Homes* domestic home energy efficiency improvement schemes move households out of fuel poverty and whether the Schemes is being successful in targeting those most in need.

Even what might on the face of it seem a simple question like what impact our *Warm Homes* interventions are having on the health of recipients was not possible in Wales before the advent of data linking. However, the ability to securely link multiple datasets in a secure setting can allow us to set our sights on examining more ambitious topics - what Professor Fiona Stanley refers to as ‘wicked problems’ – issues with multiple causes and multiple potential solutions such as substance misuse, mental health,

obesity and climate change. Only by bringing together data from multiple sources can we hope to understand the full complexity of these multifaceted, crosscutting problems and identify how we can use more joined up policy solutions to solve them.

Equally, with complex social issues where multiple services are involved in delivery, e.g. ‘missing children’, domestic violence or substance misuse, linking administrative data from multiple sources is the only way we can hope to understand the ‘journey’ of the user to inform service improvement or deliver an outcomes evaluation by identifying their ‘destinations’.

Some of the potential benefits of linking administrative data spring from the fact that the data tend to be comprehensive, covering the whole population of eligible individuals or households, so that for routine health records, for example, almost the whole population is included. This is why the Office for National Statistics have been working for many years on perhaps the most ambitious of all data linking projects - the possibility of transforming the Census by linking together administrative data supported by survey information to produce census-type outputs^{xiii}.

However, beyond the Census, the fact that some administrative sources, under the rigorously controlled conditions available in secure settings, can be linked for the whole population, suggests a range of further possibilities for improving the evidence base. For some rare events or small population groups, for example some of those with protected characteristics under the UK Equality Act 2010, linked administrative data is the only realistic way to deliver usable data. For some purposes, only a population-wide analysis will do, for example Welsh Government are currently working with the Administrative Data Research Centre for Wales on a project to examine the feasibility of constructing an individual deprivation measure in Wales using administrative data^{xiv}. One of the main limitations of the Welsh Index of Multiple Deprivation (WIMD)^{xv}, the Welsh Government’s official measure of relative deprivation in Wales, is that, because it is an area-based measure, whilst an area may be designated as being multiply deprived, all those living within a particular area will not necessarily suffer multiple, or indeed any, deprivation^{xvi}. This project therefore has the potential to help us to better understand how deprivation is distributed in Wales and therefore to support a more effective approach to targeting a range of services to individuals in society who are in greatest need.

By taking these approaches to their logical conclusion, in the long term and with significant investment of both time and funding, the use of secure settings and linked administrative data might eventually allow us to bring together data across a range of topics for the whole population, allowing us to analyse that data as we do survey data now, so that when exploring the impact of interventions, we see whether, for example, the outcomes

vary by educational attainment or socio-economic status. In theory, we could build tools within secure settings for e.g. policy options appraisal, allowing us to more accurately model the future impact of different policy options on the affected populations.

With careful planning, and ensuring we collect individual-level data on all our interventions, it would become possible to undertake a reciprocal data linking project to the *Individual Deprivation Feasibility Study* looking at the intersections between policy interventions at the individual level. With the *Supporting People*, *Flying Start* and *Fuel Poverty Data Linking Projects*, we have demonstrated that we can flag the recipients of multiple interventions. It would be useful to analyse - and to map - where and among which population groups these interventions are taking place. The ability to compare an individual's 'multiple deprivation' with the 'multiple interventions' they are receiving, as well as to examine whether improvements in outcomes are being achieved, would have the potential to drive a revolution in policymaking.

Allowing ourselves to fall back down to earth, however, the following brief UK and International examples, none of which would have been possible without the linking of anonymised data in a secure setting, illustrate the kinds of policy impacts that can already be achieved:

4.1 Suicide after release from Prison (Australia)

To study the suicide rate in the period immediately after release from prison, data was linked from the New South Wales Department of Corrective Services and the Australian Institute of Health and Welfare for a retrospective cohort of 85,203 adult offenders who had spent time in full-time custody between 1988 and 2002^{xvii}. The suicide rate in men in the two weeks after release was 3.87 (95% CI, 2.26–6.65) times higher than the rate after six months. After six months, the suicide rate approached the rate observed in custody. Similar studies have been done in Sweden^{xviii} and the US^{xix}, all concluding that prisoners are at a heightened risk of suicide in the immediate post-release period. This evidence has led to more focus by services on the first six months of the transition to life outside prison and to increased attention on prisoners with a previous suicide attempt.

4.2 The Supporting People Data Linking Project (Wales)

The Welsh Government *Supporting People* programme provides £124 million per year of housing-related support to help 60,000+ vulnerable people, some of whom have chaotic or risky lifestyles e.g. substance misuse, domestic violence, to live as independently as possible. To explore the impact on the health service use of those who accessed support, the feasibility study for this project linked *Supporting People* administrative data from two local authorities in Wales to routine health records using the secure setting of the Administrative Data Research Centre

for Wales. Those receiving support showed an average decline of 0.8 days on which GP events occurred per service user compared with an average of 0.2 days for those who were 'unsuccessful' or 'failed to engage'. This may suggest *Supporting People* was helping service users in ways that reduced the burden on the NHS.^{xx} The feasibility study having demonstrated the analysis of linked *Supporting People* data was possible and that it may be able to demonstrate cost offsetting to assist in evaluating the effectiveness of the programme, the study has now been funded for a further four years.

Welsh Government Cabinet Secretary for Communities and Children, Carl Sargeant AM, said:

"I was really pleased to learn about the Supporting People Data Linkage study. It uses innovative data-linking techniques to help us understand the impact of the Supporting People Programme on the lives of those who receive support, and how it helps them to engage with health services in more appropriate ways. This work has a lot of potential and I was pleased to be able to approve funding for it to continue to cover the whole of Wales. I look forward to seeing the results"

4.3 Participation in Higher Education (UK)

In order to examine the relationship between young people's backgrounds and their participation in higher education, Bowes et al^{xxi} linked administrative data collected by the Department for Education and the Higher Education Statistics Agency to survey data collected as part of the Avon Longitudinal Study of Parents and Children (ALSPAC) and the Longitudinal Survey of Young People in England (LSYPE). This project offered some challenges for policymakers seeking to identify interventions to help support young people to progress in their chosen pathways, concluding that the interconnected nature of factors made it impossible to determine a single barrier on which policymakers could focus to fix the problem of lack of progression. Instead, potential interventions likely to be most effective in widening access to higher education would incorporate both 'push and pull' features, addressing barriers while promoting the benefits e.g. raising awareness of the financial and non-financial benefits of higher education and the opportunities available to people with higher level qualifications amongst young people and their support networks particularly during the transition from Key Stage 3 to Key Stage 4.

4.4 The Fuel Poverty Data Linking Project (Wales)

Between 2012 and 2016, Welsh Government has invested over £217 million in its Welsh Government Warm Homes Programme, improving over 39,000 homes, reducing energy bills and helping households to heat their homes at

a more affordable cost. Linking scheme data to routine health records, the Project demonstrated a statistically significant positive impact of the Nest scheme on overall respiratory health and on asthma for recipients and suggested a protective effect on general health as measured by levels of prescribing for infections. Additionally, the results suggested a positive impact on emergency hospital admissions for both cardiovascular and respiratory conditions, although due to small numbers the effect did not reach the level of statistical significance.^{xxii} The findings have been widely welcomed by stakeholders, have been beneficial in developing relationships with health professionals and fed into policy decision making about the eligibility criteria for the successor scheme, which is currently under development.

Welsh Government Cabinet Secretary for Environment and Rural Affairs, Lesley Griffiths AM, said:

“These findings show not only is our successful Warm Homes Nest scheme helping to reduce energy bills while reducing emissions, but is also having a positive impact on the health and well-being of some of Wales’ most vulnerable households. This is why I am pleased we are continuing to invest in home energy efficiency improvements for low income houses and have committed £104million over the next four years to improve up to 25,000 more homes through the Welsh Government Warm Homes Programme”.

4.5 The Educational Outcomes of Children in Care in Manitoba (Canada)

By linking anonymous records of school performance, birth outcomes and Child and Family Services (CFS), Brownell et al^{xxiii} could for the first time paint a clearer picture of children in care and how they were doing in school. The findings allowed them to see where the trouble spots and the success stories were, looking at a variety of measures of school performance from kindergarten to high school to see if children in care had different results compared with those who weren’t in the system. They looked at all Manitoba children who had been in care for any length of time compared with those who had never been in care but whose families had received support from CFS as well as those who had no involvement with CFS. Children in care were found to be more likely to face a variety of challenges e.g. disability, poverty, mothers who started having children young or reported using alcohol or drugs during pregnancy, all of which were also associated with poorer educational outcomes. The results revealed that, compared with children who had never been involved with CFS, children in care were twice as likely not to be ready for school in one or more areas, were less likely to perform well in maths and reading and were less likely to graduate from high school. As a result, the Manitoba government announced it was revising the funding model

of CFS to expand support for families with the goal of keeping children in their homes and communities rather than placing them in care.

5. Conclusion: it would be a ‘Failure of Duty’ by Government not to link data

If the arguments and examples outlined above haven’t yet convinced you of the value of safe settings for policy, it’s worth pointing out that a growing number of researchers are starting to examine the risks of not using data and to argue that it’s unethical or a ‘failure of duty’ by government not to allow administrative data to be anonymised and linked, generating research findings to better inform their decision making. Cukier and Mayer-Schoenberger, although their observations were made in the context of the broader category of ‘big data’, argue that it will profoundly change how governments work and ‘when it comes to .. providing public services .. those who can harness big data effectively will enjoy a significant edge over others’. Furthermore, they advise us to ‘shed our preference for highly curated and pristine data and instead accept messiness: in an increasing number of situations, a bit of inaccuracy can be tolerated, because the benefits of using vastly more data of variable quality outweigh the costs of using smaller amounts of very exact data.’

There are most certainly significant challenges that need to be overcome before we, as government, can get the best out of the growing secure data infrastructure of the UK. As Holman et al point out, the ‘creation of a data linkage system ... demands leadership, interagency and inter-sectoral cooperation, a dedicated group of users who drive reforms, and perseverance’. However, they also point out that the data linkage system they have created in Western Australia has made unbiased contributions to medical knowledge and identifiable advances in population health. The message from such leading lights in the world of data linkage as Professor Fiona Stanley is that ‘data linking saves lives’; she argues: ‘[“wicked problems”] don’t only need good data, they need data to be joined up’. Where the development of best practice protocols and procedures to address privacy concerns is a fundamental first step on the path to giving government agencies confidence in sharing data, thereby opening up all the avenues outlined above and potentially gaining all the benefits linked data can offer, the value of safe settings to evidence-based policymaking simply cannot be underestimated.

References

- ⁱ Desai, T., (2017) Disseminate Access Not Data; Ritchie, F., (2017) ‘5 safes’ as a model for data access decision making; Welpton, R., (2017) Skills Gaps: all Data for Policy 2017
- ⁱⁱ <https://statswales.gov.wales/Catalogue>
- ⁱⁱⁱ <http://lle.gov.wales/home>
- ^{iv} <https://saildatabank.com/>
- ^v Cameron, D., Pope, S., Clemence, M. (2014) Exploring the public’s views on using administrative data for research purposes,

Ipsos MORI Social Research Institute

<http://www.esrc.ac.uk/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>

^{vi} IPSOS Mori (2014), The One-Way Mirror: Public attitudes to commercial access to health data,

<https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>

^{vii} Royal Statistical Society (2014), Royal Statistical Society research on trust in data and attitudes toward data use / data sharing, <http://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf>

^{viii} CLOSER = Cohort and Longitudinal Studies Enhancement Resources

^{ix} Allison Park Blog: Digital Economy Bill: how academic research + government data = a rich mine of information <https://ioelondonblog.wordpress.com/2016/07/28/digital-economy-bill-how-academic-research-government-data-a-rich-mine-of-information/>

^x Care.data: How did it go so wrong?

<http://www.bbc.co.uk/news/health-26259101>

^{xi} Interview with Prof Fiona Stanley, ABC News, Australia:

<http://www.abc.net.au/worldtoday/content/2013/s3739795.htm>

^{xii} <https://www.researchgate.net/project/Welsh-Government-ESRC-PhD-An-Investigation-into-the-Impact-of-Disability-on-Employment-in-the-UK>

^{xiii} <https://blog.ons.gov.uk/2017/06/20/an-administrative-data-census-transforming-the-census-for-the-future/>

^{xiv} <https://www.researchgate.net/project/The-Welsh-Government-Individual-Deprivation-Data-Linkage-Feasibility-Project?updateId=59846f49b53d2ff30bd832f4&pua=true>

^{xv} <http://gov.wales/statistics-and-research/welsh-index-multiple-deprivation/?lang=en>

^{xvi} Deas, I., Robson, B., Wong, C. and Bradford, M. (2003) Measuring neighbourhood deprivation: a critique of the Index of Multiple Deprivation. *Environment and Planning C: government and Policy*. 21(6):883-903

^{xvii} Kariminia A, Law MG, Butler TG, Levy MH, Corben SP, Kaldor JM, Grant L. (2007) Suicide risk among recently released prisoners in New South Wales, Australia. *Med J Aust* 2007; 187 (7): 387-390.

^{xviii} Haglund A, Tidemalm D, Jokinen J, Långström N, Lichtenstein P, Fazel S, Runeson BS. *J. Clin. Psychiatry* 2014; 75(10): 1047-1053. DOI: 10.4088/JCP.13m08967.

^{xix} Binswanger IA, Blatchford PJ, Lindsay RG, et al. Risk factors for all-cause, overdose and early deaths after release from prison in Washington state. *Drug Alcohol Depend.* 2011;117(1):1–6. DOI: 10.1016/j.drugalcdep.2010.11.029

^{xx} The Supporting People Data Linking Project:

<http://gov.wales/statistics-and-research/supporting-people-data-linking-feasibility-study/?lang=en>

^{xxi}

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474269/BIS-15-462-understanding-progression-into-higher-education-final.pdf

^{xxii} The Fuel Poverty Data Linking Project:

<http://gov.wales/statistics-and-research/fuel-poverty-data-linking-project/?lang=en>

^{xxiii} Brownell, M. Chartier, M. Au, W. MacWilliam, L., Schultz, J. Guenette, W. and Valdivia, J (2015) *Manitoba Can Do Better For*

Kids In Care: a summary of the report, The Educational Outcomes of Children in Care in Manitoba. Manitoba Centre for Health Policy